DOCUMENT RESUME

ED 209 365 OD 021 695

AUTHOR Renwick, George W.
TITLE Evaluation Handbook for Cross-Cultural Training and Multicultural Education.
INSTITUTION Intercultural Network, Inc., La Grange Park, Ill.
REPORT NO ISBN-0-933662-08-4
PUB DATE 79
NOTE 64p.; Not available in paper copy due to institution's restriction.
AVAILABLE FROM Intercultural Press, Inc., 70 West Hubbard Street, Chicago, IL 60610 ($4.50; discount on bulk orders).

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS Cross Cultural Training; *Evaluation Criteria; *Evaluation Methods; Measurement; *Multicultural Education; *Outcomes of Education

ABSTRACT

The process of evaluating multicultural education programs with respect to goals, planning, implementation and effects is examined in this handbook. The roles of evaluation participants are defined, and the content of what is to be measured is classified according to five major categories of possible effects upon the learners: knowledge, perceptions, attitudes, skills and patterns of behavior. Also examined are specific methods of measurement which include paper and pencil methods, discussion, production and observation, and appropriate procedures for tabulating, recording, and interpreting the results. (JCD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED209365

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Margaret D. Pusch
Intercultural Press, Inc.

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

evaluation handbook

For Cross-Cultural Training and Multicultural Education

GEORGE W. RENWICK

Intercultural Network, Inc.
LaGrange Park, Illinois

US021645

INTERCULTURAL NETWORK, INC.
906 N. Spring Avenue
La Grange Park, ILL. 60525

This book was originally published as part of a larger work, MULTICULTURAL EDUCATION: THE CROSS-CULTURAL TRAINING APPROACH, which may also be ordered from the Intercultural Network, Inc.

ISBN NO. 0-933662-08-4
Library of Congress Catalogue Number 79-92378

© Intercultural Network, Inc.
Printed in the United States of America

AUG 5 1981

CONTENTS

	Page
INTRODUCTION	1
1. WHO DOES THE EVALUATION	3
2. FOR WHOM THE EVALUATION IS DONE	3
The Students and Program Participants	4
The Teacher-Trainer	4
Administrators and Managers	4
Colleagues	5
Outsiders	5
3. WHEN THE EVALUATION IS DONE	6
4. WHAT IS TO BE MEASURED	9
Knowledge	11
Perceptions	12
Attitudes	12
Skills	13
Patterns of Behavior	14
5. HOW THE MEASUREMENT IS TO BE DONE	16
Paper and Pencil Methods	18
A. Check Lists	18
B. Rank Ordering	19
C. Scales	21
D. Written Training Exercises	23
E. Critical Incidents	24
F. Case Studies, Films, Photographs	25
G. Sentence Completion	27
H. Questions	28
I. Journals	32
Discussion	33
Enactment	34
Production	35
Observation	36
Unobtrusive Measures	37

6. HOW THE TABULATION OF THE RESULTS CAN BE ENJOYED..... 40

Check Lists..... 41
Scales..... 41
Multiple Choice Questions..... 42
Essay Responses and Short Answers..... 42

7. APPROPRIATE FORMAT FOR THE FINDINGS..... 47

Check Lists..... 48
Scales..... 49
Essay Responses and Observations..... 50

8. INTERPRETATION OF THE FINDINGS..... 51

9. CONCLUSIONS..... 53

RESOURCES..... 57

INTRODUCTION

Our most fundamental purpose, as teachers and trainers responsible for intercultural courses and programs, is to stimulate learning. Our own professional preparation, the design of our courses and programs, the particular methods and materials we carefully choose--all these are intended to deepen the learning of our students and participants. Seldom, however, do teachers and trainers ask themselves, has there actually been any learning? If so, what kind of learning? Is it what we intended? If our best efforts have resulted in some learning, how much learning has there been? For whom? And how long will the learning be retained? Does it become the basis and means for still further learning over time, or is it quickly extinguished?

As is true with most questions in the social sciences, it is difficult to come up with conclusive answers. When dealing with people, especially changes within and between people, proof is elusive. Evidence, however, is available; indications, clues and signals can be obtained; they suggest answers to these fundamental questions and are therefore indispensable.

Evidence is gathered through evaluation. Revealing evaluation does require some thought and effort, but it can be done with less technical knowledge than many teachers and trainers realize. For some kinds of evaluation, for example, a knowledge of statistics is not necessary. A rather simple, carefully designed evaluation of a carefully designed course or program can yield surprising insights and suggest new, very promising directions. Evaluation can become, in other words, the primary channel through which the teacher learns. What is learned through this channel is often extremely interesting. Further, of course, what the teacher learns enables the teacher to increase what the student learns.

Let's be more specific. We are not concerned with all kinds of learning; we are concerned, much concerned, about learning how to function and grow within, and contribute to, pluralistic societies. Multicultural education and cross-cultural training can be defined as

a structured process designed to foster understanding, acceptance and effective interaction among people of many different cultures. The question we must now ask is: How do we know, in the midst of this process, whether we are actually fostering understanding, acceptance and effective interaction?

Still more specifically, through multicultural education and cross-cultural training we attempt to build an awareness of one's own cultural heritage, and to build those skills in analysis and communication that contribute to an ability to function effectively in multicultural environments. The question now: How do we know that students and program participants, as a result of our efforts, have developed more cultural self-awareness and more competence (and confidence) in intercultural analysis and intercultural communication? We know this, or can at least estimate this, through basic evaluation procedures.

Fortunately, there is now much work in the field of evaluation to draw upon. During recent years, as the number of institutions and projects competing for limited financial resources has rapidly increased, the requirements for systematic evaluation have become more frequent and more stringent. Project directors, educators and others are now held more accountable for what they do (and fail to do). More attention has therefore been given to the purposes and especially the methods of evaluation. Numerous organizations and several journals specializing in evaluation have been established (some of these are listed with the resources at the end of this handbook). Evaluation has now become a distinct and very lively field.

Our purpose here is to explain some of the basic procedures of sound evaluation and some of the methods of measurement which are now being used. We have selected those which are most helpful to those involved in cross-cultural training and multicultural education, and especially helpful to teachers and trainers who are experimenting with new cross-cultural methods and materials.

How, then, does the teacher or the trainer become a responsible evaluator? The first step, of course, is to recognize the importance of evaluation to oneself and one's students or program participants, and to commit oneself to doing evaluation. Next, a number of basic questions must be considered:

1. Who does the evaluation?
2. For whom is it done?
3. When is it done?
4. What is to be measured?
5. How is the measurement to be done?
6. How can the tabulation of results be enjoyable?
7. What is the appropriate format for the findings?
8. How are the findings to be interpreted and used?

We will consider each of these questions in this handbook and will outline some of the alternatives within each.

1

WHO DOES THE EVALUATION

There are a number of reasons why it is sometimes best to have the evaluation done by someone who is not involved in the course or program, someone in another part of the school or the organization, or perhaps better, someone from outside. Such a person brings objectivity to the task and may bring special expertise. There are several organizations which evaluate intercultural teaching and training activities; information on these is included in the list of resources.

There are also reasons, however, why it may be best for teachers and trainers to do the evaluation themselves. Although it may be difficult for them to be objective, they are certainly the ones most thoroughly acquainted with the goals, content and methods of the courses they teach and the programs they conduct, as well as the needs, learning styles and aspirations of the students and program participants. Furthermore, there is seldom time or funding to enlist someone else. If there is to be an evaluation, therefore, there is normally little room for choice. It is usually the teacher or trainer who must do it. In the remaining areas, however, there is a great deal of room for choice.

2

FOR WHOM THE EVALUATION IS DONE

This decision will influence significantly all subsequent decisions relative to evaluation. The choice of the "consumer" of the evaluation - the person or group for whom it is being done - will influence the purpose of the evaluation, when it is done, what is measured, how the measurement is done and how the findings are used.

There are several possible consumers. At the beginning of an evaluation effort, the evaluator should choose the one to which priority will be given:

THE STUDENTS AND PROGRAM PARTICIPANTS: In some of their evaluation efforts, teachers and trainers will probably want the findings, and in fact the evaluation process itself, to be of primary and most direct benefit to their students and participants. In this case, the evaluation is designed to provide pertinent information to them for their consideration and discussion, or to give them periodic check points and readings on their progress.

Evaluation designed primarily for the student or participants can serve another important purpose, perhaps the most important purpose. It can enable them, through the questions it asks or thought it stimulates, to continue to evaluate their learning after the formal evaluation in a particular course or program is completed. Through their structured practice in the classroom, students and participants can learn habits of constructive self-evaluation which will be used and developed further during later education and later experiences in life. Of greatest benefit to them, therefore, may be the process of the evaluation which serves as a stimulus to further learning. The design of this process is therefore extremely important, both for the information it yields and the patterns it imparts.

THE TEACHER AND THE TRAINER: Certain times during the year (or during their careers), teachers and trainers may want or need a reading on themselves and on their courses or programs. Evaluation can provide them with information on how they come across and how the subject matter they choose and the methods they use are being received. The end in mind is to increase their professional growth and effectiveness and to enable them to modify where indicated some of their strategies.

All conscientious teachers and trainers, of course, try to pick up indications of learners' responses (and others' responses) every day. A more systematic procedure, however, can be more accurate and can be a very valuable supplement. Because the person to gain most directly from this kind of evaluation are teachers and trainers themselves, the findings are often not given to others (students, administrators, etc.).

ADMINISTRATORS AND MANAGERS: In their attempts to determine the quality of their school's curriculum, administrators may periodically ask teachers for reports on their intercultural courses. Or an administrator may call upon them to justify the continuation of a course, the introduction of a new course or special intensive program, or the introduction of a major new approach within existing course. In each case, the evaluation is done primarily for the administrator. The administrator's interests and questions must therefore be the ones clarified at the outset and addressed in the evaluation design.

Similarly, managers of business and other organizations who request or simply approve a cross-cultural training program usually want to know the consequences of the program. In obtaining the

evaluation, therefore, the trainer must give priority to the questions and concerns of the manager.

COLLEAGUES: Teachers and trainers often work quite isolated from one another. The approaches and strategies of each, and the consequences of these for the students or participants, are seldom known to one's colleagues. This lack of communication, of course, limits the learning of everyone in a field and retards the development of multicultural education and cross-cultural training.

Both teacher and students, as well as trainers and participants, stand to benefit from the experiences and experiments of others. Teachers and trainers can begin to provide some of this benefit by conducting careful evaluation designed primarily to assist themselves and their colleagues in their common endeavor and to advance their developing field. Such evaluation efforts are usually most helpful if they are designed cooperatively by several colleagues. The evaluation can then be conducted in several classes and programs, the findings can be jointly analyzed and the conclusions jointly drawn. Perhaps, occasionally, articles can then be written. Evaluation of courses and programs, and cooperative analysis of their effects, are especially important where new intercultural techniques are being tried, modified and integrated with regular curricula and training sequences.

OUTSIDERS: Teachers and administrators, trainers and managers, sometimes find it advantageous (or even necessary) to describe their courses and programs in a clear, convincing way to particular individuals and institutions outside their schools and organizations. These would include state and federal government agencies, accreditation units, potential funding sources, professional societies, school boards, other schools, other organizations and potential clients. If the teachers and trainers are to represent their efforts in a credible manner, they must base their presentations on sound evaluation. In such cases, it is the requirements of this outside party that must be kept in mind while designing the evaluation. The subsequent choices will be shaped accordingly.

The basic purpose of evaluation in academic and professional education, then, is to provide for learners, evaluators, administrators, colleagues and outside parties the specific information on courses and programs which will be directly related to their concerns and immediately useful in their work.

In determining what will be most useful to the one to whom priority has geen given, it is helfpul to look especially at the decisions the person or group must make. Once their major decisions have been clarified, the evaluator will be able to specify what information is necessary in order to help them make a sound decision. Knowing the kind of information needed, the evaluator is then better able to make subsequent choices regarding when to do the evaluation, what to measure, the methods to be used and the format in which the findings are to be presented.

3

WHEN THE EVALUATION IS DONE

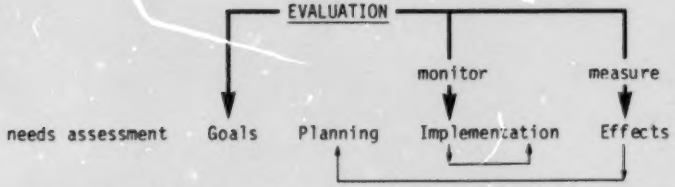
Courses and programs are seldom evaluated at all. When they are, the evaluation is usually looked upon as a means for finding out generally how things went and is tacked on at the end.

Evaluation, however, is also a very valuable means for finding out how things <u>are going.</u> It should therefore be done <u>during</u> the course or program.

In addition, evaluation is the most useful means for finding out how things <u>should go</u>. It is therefore important to begin asking questions about the evaluation, using evaluation methods (and studying the findings from previous evaluations) <u>before</u> the course or program, that is, during the planning phase.

In order to contribute all it is capable of, evaluation must be built into the entire teaching and training process and should inform each phase of the process. It is, in fact, a primary means through which the different phases of a course or program can be integrated with one another. Next to the teacher or trainer himself, evaluation can be the major source of course integrity.

A diagram is presented below. It clarifies the purposes of evaluation and identifies the phases of the teaching and training process at which evaluation can be most productive.



As illustrated in the diagram, the first step in preparing a course or program is the needs assessment. The specific needs of students and of participants in training programs, if estimated at all, are often estimated on the basis of hunches. In non-technical learning (involving, for example, cultural self-awareness and communicative competence), as in technical learning, systematic assessment is essential in ensuring that the course or program meets the real needs of the learners. It is upon the precise definition of their needs that the goals of the learning sequence are formulated and the sequence itself is designed. Also, it is (partially) in terms of students' or participants' specific needs that the goals, implementation and effects of the sequence are later evaluated.

It is helpful to think of assessment and evaluation as somewhat different activities. Assessment looks more at given conditions and then describes those conditions. Evaluation looks more at processes and then determines their worth and their consequences. The procedures and methods for assessment and evaluation, however, are often the same. In preparing their needs assessment, therefore, the teachers and trainers can choose from the alternatives outlined later those which are most appropriate to their particular students and to their particular purposes. The needs assessment phase is usually necessary whether the evaluation is to be done primarily for the students or participants, the teacher or trainer, an administrator or manager, colleagues or an outside party.

Looking now at the rest of the diagram, the reader can see that it assumes the evaluation is to be used primarily by the teacher or trainer. The monitoring function of evaluation, for example, feeds back immediately into the course itself in order that the teacher can make the indicated adjustments as the course goes along. Similarly, the measuring function does not feed directly into the discussions of the learners or the decisions of an administrator or a manager (although, of course, with some different choices it could). Rather, the measuring function feeds back into the planning phase for which the teacher or trainer is responsible and during which fundamental decisions must be made.

The effects which are identified and measured in the diagram are those which are evident at the close of the course or program. The longer-term effects, those which are measured during a follow-up evaluation, are not represented. Needless to say, however, they are crucial. A follow-up evaluation is seldom done but is usually essential if accurate conclusions are to be drawn regarding the impact of

the teaching or training effort--and if confident planning is to be done for the next one. The follow-up evaluation can be done one month, three months, six months or longer after the course or program, or it can be done at each of these points (thereby tracing multiplying or diminishing effects) Care must be taken in interpreting the findings, of course, because there are inevitably many other factors during the post-course period which can influence students' and participants' learning and therefore their responses during the follow-up evaluation (this influence is mentioned again in our discussion of methods of measurement). The findings from the follow-up evaluation, as with the findings at the close of the course, feed into the planning phase for the next course.

4

WHAT IS TO BE MEASURED

As with the choices in each area, the choices regarding what one should measure will be influenced by the choices which have already been made in other areas. If the findings are primarly for the students, for example, what is measured will sometimes be different from what would be measured for an administrator. Similarly, the choices made in this area will influence choices made in subsequent areas; what is to be measured influences especially the next area: how the measurement is to be done.

The choices here, as elsewhere, are also influenced by the culture and preferences of the teacher, and especially by the cultures and preferences of the students. Investigation of certain subjects is inappropriate for people from some cultures. Investigation of other subjects is not only appropriate but expected. Fundamental cultural characteristics must therefore be kept in mind during the design (and the introduction and administration) of an evaluation, not only out of consideration for the people but also out of concern for the accuracy of their responses. Their responses to a question which is inappropriate, for example, may reflect their resentment toward the question itself (or toward the questioner) rather than their judgment (which is actually being sought).

What, then is to be measured? Lets take teachers and their intercultural courses as an example. There are three basic alternatives: we can measure the teachers, their courses, or the consequences of the teachers and their courses - the effects upon the learners. Once this choice has been made, there are further alternatives to consider. With regard first to the teachers, one might want to look at their knowledge of their subject, their control over their classes, their ability to interest and motivate students, and so forth. If one is working with multicultural classes, especially if the teachers are trying to develop the students' abilities as intercultural communicators, an evaluation might measure the teachers'

abilities as role models: their own cultural self-awareness, their appreciation of the contrasting values, goals and learning styles of the students, their ability to act as cultural interpreters or brokers, their ability to identify the cultural sources of conflict among the students and deal creatively with these, or their capacity to communicate their ideas (and perhaps feelings) to all of their students in a clear, receivable manner.

With regard to the course or program, if this is to be the focus of the evaluation, one might look at the goals, components, sequencing (does one component build upon another?), pace, atmosphere, organization, facilities, and so forth. If there are a number of cultures represented in the course or program, one of the questions an evaluator might ask is this: Are the demands which the structure and strategies make upon the learners (for participation, initiation or self-disclosure, for example) compatible with their customary ways of learning and relating in formal settings?

With regard now to the effects upon the learners, it is with these that this and the next area are primarily concerned. Measurement of effects sometimes takes more thought, and it is in specific effects that teachers and trainers in multicultural education are usually most interested. Furthermore, a careful study of the effects of a course or program will produce findings which, at least indirectly, provide some critique of both the teacher and the course.

A course or program has effects upon both the individual learner and upon the group of learners as a whole. We will outline first some of the possible effects upon the individuals, then the effects upon the group.

There are five major categories of possible effects upon the individual. An evaluation of effects measures changes which occur within one or more of them. The categories are:

1. knowledge
2. perceptions
3. attitudes
4. skills
5. patterns of behavior

Each of these, of course, is related to the others. Each is integrated with the others in the individual student and participant. Looking at the way a student handles intercultural conflict, for example, we might find the following: the student has studied and learned about the fundamental characteristics of both his own culture and the culture of the person with whom he is now in conflict (knowledge). Given this knowledge, the student's images of the other person, especially his images of the other's background, are rather accurate perceptions. Because of his knowledge and his perceptions, the student may be open toward the other person, willing to deal with the conflict and confident as he begins to do so (attitudes). Given his knowledge, perceptions and attitudes, the student may be able to explain his own position in a way that the other person can understand, and may be able to analyze for the other person the cultural differences which caused or contributed to their conflict

(skills). As the student explains, analyzes, and interprets with this person and then with others, as he conveys his attitudes to this person and then to others, and as he responds to others, he practices and develops new modes of relating and communicating (patterns of behavior).

If the student or participant is in a multicultural setting, of course, he has an excellent opportunity to gain his knowledge of other cultures directly from representatives of those cultures. Given that cross-cultural training is designed to take maximum advantage of this opportunity through the structured interaction, the individual's knowledge about other cultures and persons is significantly enhanced--and his perceptions and attitudes are often shaped accordingly.

Intercultural courses and training programs are also designed to enable persons from contrasting cultures to learn about the process of perceiving -- about the process itself -- and about the process of communicating; the student is then given opportunities to engage in both. Intercultural approaches, therefore, contribute to an additional, and fundamental, kind of knowledge -- a knowledge of process (which can be measured), as well as contributing to the development of specific perceptual and communication skills (which can also be measured).

A word about feelings. A variety of feelings emerge during an intercultural course or program, especially feelings among the students for the teacher and for the other students. These feelings may be some of the most significant effects of the course: affection between an Anglo and a Chicano, rapport between a Black and a White, frustration between an Arab and a Chinese, and so forth. Such feelings are usually not addressed directly in an evaluation. The reasons are that direct investigation of feelings is sometimes inappropriate, and feelings are assumed to be reflected in attitudes and patterns of behavior, both of which are usually more easily (and more appropriately) measured.

Carefully designed and integrated intercultural teaching and training strategies, therefore, will probably produce all five measurable effects -- some more than others, perhaps, but all five will usually be present. The question is, Which of the five should be measured? It is probably best if teachers and trainers begin their evaluation efforts by choosing just one, and perhaps just a few of the alternatives within that one. Let's look at some of the alternatives within each. The reader will undoubtedly think of (and perhaps choose) others.

KNOWLEDGE: If it is knowledge that is chosen, one might want to examine some of the following and may want to ascertain any changes in these as a result of the course or program. These include the learner's knowledge of:

1. His own culture, his own distinctive values, goals and patterns of learning and behaving.
2. Other cultures studied or represented in the class (Puerto Rican, Nigerian, Mexican, etc); in other words, his "cultural literacy."
3. The influence of culture upon personality and behavior.
4. The process of perception.
5. The process of communication.
6. The sources of intercultural conflict.
7. Stresses unique to, or augmented in, multicultural environments.
8. Culture shock, culture fatigue, role shock and role fatigue, and what to do about them.
9. His own resources to be mobilized when he is isolated in a new environment.

PERCEPTIONS: Courses and programs also influence perceptions. Perceptions are sometimes, in fact, the major focus of a course or selected strategies within a course. The evaluator may want to get a reading on the learners' perceptions and trace changes in them. An evaluation designed to do this might measure the learners':

1. Images of themselves.
2. Images of people from their own culture.
3. Images of people from other cultures studied or represented in the class or training group (Chinese, Saudis, Vietnamese, English, etc.).
4. Perceptions (attributions) of the causes of their own and others' behavior.
5. Perceptions of the intentions of culturally different persons toward them.

ATTITUDES: The teacher or trainer may be primarily interested in the effect of the course or program upon the learners' attitudes. As an evaluator, then, he or she should measure some of the learners' attitudes toward:

1. Themselves (cultural self-affirmation, self-esteem).
2. People from their culture.
3. Strangers in general (ethnocentrism).
4. People from another particular culture.
5. Initiating contact with people from different cultures.
6. Assisting individuals from a different culture.
7. Being assisted by individuals from a different culture.
8. Being involved in a close relationship with someone from a different culture (a Black, an Italian, a Native American, etc.).

Other possible attitudes include the learners':

9. Curiosity about different cultures, their values, priorities, lifestyles, etc.
10. Motivation to pursue relationships with individuals from different cultures.

11. Confidence before and during intercultural interaction.
12. Comfort during, and enjoyment of, interaction with individuals and groups which are distinctly different.

SKILLS: Most teachers and trainers want their efforts to have practical consequences. They want to develop, in those who would learn from them, certain skills and abilities which will result in clear understanding and constructive behavior during inter- cultural interaction. An evaluation designed specifically to measure such skills and abilities would concentrate on one or a few of them. In making a choice, the teacher and trainer should consider the learners' ability to do the following:

1. Read, write and especially speak the appropriate second language. (Because of the number of methods and tests available for measuring language ability, this chapter does not consider this kind of evaluation.)
2. Learn about their own and other cultures (information gathering skills).
3. Analyze systematically basic characteristics of their own culture, especially those which influence most directly their interaction with individuals and groups from different cultures.
4. Learn from individuals from other cultures.
5. Explore the basic characteristics of a different culture, especially those which influence most directly the interaction of people from that culture with them.
6. Analyze quickly and accurately a given intercultural exchange.
7. Understand both the words and the meaning of a person during an intercultural exchange.
8. Withhold judgment while getting to know an individual or group from a different culture.
9. See and sustain conflicting viewpoints.
10. Tolerate ambiguity and ambivalence (in themselves, in others and in relationships).
11. Anticipate specific points of potential difficulty in intercultural relations.
12. Recognize and clarify value, perceptual and other culturally-based conflicts.
13. Put themselves in the position of someone from a different culture.
14. Accept, respect and accommodate major cultural differences.
15. Find common ground and common objectives with someone whose values and preferences are, for the most part, quite different from their own.
16. Motivate an individual or group whose priorities are quite different from their own.
17. Adjust their images of individuals and groups from different cultures so as to make them more accurate (cognitive flexibility vs. rigidity).
18. Analyze accurately and respond constructively to prejudice and discrimination against them.
19. Cope with the particular stresses inherent in multicultural environments.

PATTERNS OF BEHAVIOR: The students' and participants' knowledge, perceptions, attitudes and skills will, to some extent, result in particular patterns of behavior. Conversely, their patterns of behavior will, to some extent, result in particular skills, attitudes, perceptions and knowledge. Some courses and programs, therefore, are designed primarily to bring about specific, directy observable, patterns of behavior. Because behavior can often be directly observed, it is often more easily measured. Teachers and trainers, therefore, may want to design their evaluation in such a way that it will locate, record and measure specified patterns of behavior in and between individual students. Some of the patterns for which they might look are the following:

1. Listening "actively" with individuals from contrasting cultures.
2. Giving constructive feedback to someone from a different culture.
3. Receiving feedback, perhaps given in an unfamiliar way, from someone from an unfamiliar culture.
4. Checking their own perceptions during an intercultural exchange with an individual or group.
5. Clarifying their own and others' intentions during intercultural interaction.
6. Shifting their behavior when one pattern is no longer appropriate to the situation or the person (behavioral flexibility vs. rigidity).
7. Taking specific, appropriate steps to reduce intercultural conflict.
8. Negotiating the misunderstandings and tensions inherent in intercultural interaction.
9. Taking specific, constructive action in the midst of culture fatigue or role shock.

We have been considering in the above an evaluation of the effects upon individual students and participants. As mentioned earlier, courses and programs also, of course, have definite effects upon groups. The particular groups with which we are primarily concerned here is the teacher with his (or her) students who comprise a given class, and the trainer with certain participants in a given program. The effects upon the group can be quite significant if the members represent contrasting cultures. Some measurement of these effects is often important, especially if the approach being used with the group is experientially based, as is much intercultural education and training. Listed below are some of the characteristics of a group which may be directly affected by a course or program, especially if it is a multicultural group. Teachers and trainers should choose particular ones which they want to anticipate, design for, work toward and then (of course) measure (perhaps at different points):

1. Level of trust between different sub-groups and within the class or program group as a whole.
2. Patterns of communication (e.g., who is included and excluded, by whom and why; development of vocabulary and non-verbal expressions unique to the group).
3. Extent to which the group agrees on common objectives.

4. Clarity with which their objectives are defined and the degree to which the students or participants are included in the defining process.
5. Emergence of leaders; extent to which support for them, cooperation with them, divides along cultural lines.
6. Development or dissolution of cliques; their number and solidarity.
7. Degree of discrimination against one sub-group by another.
8. Degree of conflict and the lines along which the conflicts are drawn.
9. Level of tension and suspicion.
10. How the group handles a crisis or challenge.
11. Degree of competition within the group, and between the group and other groups.
12. The extent to which the group contributes to other groups, school projects, organization-wide efforts, etc.
13. The extent to which the members of the group have a sense of group identity.

We have now considered a variety of things which the teacher and the trainer may want to evaluate: the teacher or trainer himself, the course or program, and the consequences of each -- the effects (upon both the individual and the group). The effects chosen should become the teachers' and the trainers' general objectives and the objectives of the course or program. These then shape the design of the course or program and become a primary basis for the evaluation.

5

HOW THE MEASUREMENT IS TO BE DONE

How the measuring is actually done, of course, depends upon the choice of subject, when it occurs and what is being measured. Some methods, for example, are suitable when measuring attitudes whereas others are suitable when measuring specific patterns of behavior. The reader may select from the alternatives below the particular methods of data collection and measurement which are most consistent with his or her situation and preference.

Several general principles should be considered as the selection is made. The first, of course, is that the methods must be as compatible as possible with the values, preferences and customary modes of response of the students and participants. The effect of certain kinds of questions upon the learners, for example, given their cultural backgrounds, should be anticipated. Some questions may be appropriate only if written, others only if oral. How the questions are worded can influence significantly how they are answered. Perhaps no direct questions at all should be asked.

Responses of any kind -- written, oral or enacted -- are influenced by the extent to which the student, for example, wants to please, compliment, show respect for or "level with" the evaluator and the other students. What is socially desirable for students from one culture is likely to be different for students from another. Attitudes toward candor, politeness, cautiousness and critiquing those in authority differ, as do attitudes toward expressing oneself or performing in front of others. The feelings of students about being rated by their peers (as well as by their teacher) are deeper for some students than for others, depending in part on the cultures from which they come. Anonymity is very important to some when they are expressing reactions and criticisms; others strongly want their names associated with their opinions.

Because modes of responding differ (as do modes of learning), a variety of methods, carefully selected and limited, is often best for a multicultural class or program, especially if the cultures represented are quite diverse and the evaluation is to be a rather thorough one.

Clarity is essential in any evaluation, but it becomes particularly important when the respondents come from different cultural, educational and language backgrounds. The format, instructions and composition of the evaluation must be very precise, consistent and simple. Examples and demonstrations of what is expected are often helpful for the students or participants, as is an opportunity to ask questions regarding procedure, meaning and so forth. Clarity and consistency are critical when the evaluation requires translation (as, of course, are equivalence and accuracy); pretesting and practicing, which are important for all methods, are especially important where they require translation or bilingual interpretation.

In addition to the cultural compatibility of particular methods, the teacher and trainer should carefully consider their effect upon the students' and participants' learning. Methods of data collection often affect directly and deeply the outcomes of a course or training program -- sometimes very positively, sometimes negatively. As teachers and trainers build their methods into their course or program design, care must be taken to ensure that the methods are consistent with the purposes of the course or program and complementary to the other components in the design. The thinking, acting and interacting required of students, for example, in order to come up with the desired data can, if the methods are chosen properly, contribute substantially to the students' learning. The needs assessment helps to establish expectations; monitoring during the course can increase motiviation by demonstrating teacher concern and responsiveness; evaluation at the close can serve as an opportunity for systematic review and valuable integration; the follow-up evaluation can provide necessary reinforcement.

Most of the methods below can be more revealing, and the results they yield more credible, if they are applied to both the class or program group in which experimentation is being done and to another (and similar)control group. The point here, of course, is to compare the results from the two groups. If changes are identified in the group in which new intercultural techniques are used, for example, but the same changes do not occur in the other group (in which the techniques were not used), then the teacher or trainer can conclude with more confidence that the changes are the result of the techniques and not the result of other factors. When the use of control groups is not possible, as is often the case in multicultural education and training, more thorough evaluation is called for and more caution is necessary in interpreting the results.

If the teacher or trainer has decided to measure a number of effects, or if he is measuring just one but has chosen a number of methods, it is sometimes advantageous to divide a large class or program group into several groups and measure just one effect or use

just one method with each group. Each group, of course, must be as representative as possible of the larger group if inferences are to be drawn regarding the larger group.

In choosing and using the method(s), it is essential that the teacher and trainer be very clear as to what exactly he or she is looking for. Every method must be selected and tailored to focus on the effects to which priority has been given. One helpful way to ensure clarity and pertinence of data is to ask oneself, "Given the purposes, format and content of this course (or program), and given the particular individuals in it, what specific questions do I want to be able to answer through my evaluation?" Once the questions are formulated (and preferably written down) choosing the appropriate methods will be easier.

Once the methods have been chosen, the responses of the students and participants will be easier, probably more on target and probably more eagerly given, if the teacher or trainer explains why the evaluation is being done, how the results are to be used and who is to benefit from their joint efforts. As is often true in intercultural communication, we assist others in responding to us if we let them know "where we are coming from."

We have suggested a number of general principles: methods must be compatible with the cultures represented, a variety of methods are often more appropriate and more productive than a single one, precision and simplicity are essential, methods of evaluation enhance or hinder learning, control groups are desirable although not necessary, the teachers and trainers must be clear as to their own objectives in doing the evaluation and can encourage the cooperation of students and participants by explaining the objectives to them. If teachers and trainers keep these principles in mind, and keep their own students and participants in mind, they can select from the following alternatives the methods that will serve their purposes most fully. An example or two is given with each alternative; each example is preceded by an asterisk(*). The examples are only suggestive, of course; teachers and trainers should construct their own questions.

PAPER AND PENCIL (Self-Report)

Methods of this kind usually require students and participants to reflect upon and respond to items which are written down, recording their responses on the same or separate sheets. The respondent usually does this alone. The results are analyzed later. The alternatives include the following:

A. Check Lists: These are usually lists of single words (often adjectives or short descriptive phrases); the respondent simply checks the particular ones which apply. This method can be used for numerous purposes, one of which is to obtain a reading of the respondents' perceptions of their own and other culture groups:

*Generally speaking, Chicanos are:
(please check as many as apply):

_____ warm	_____ aggressive
_____ critical	_____ tolerant
_____ confident	_____ dishonest
_____ insensitive	_____ generous
_____ sincere	_____ superficial
_____ cold	_____ respectful
_____ friendly	_____ idealistic
	etc.

This kind of question can become the basis for evaluation by asking it before the course or program, at the close and perhaps again some months later. If the course or program has brought about changes in learners' perceptions, some of these would probably be indicated by the different adjectives (or short descriptive phrases) chosen at the different points. For further consideration of this method, see the section on the Adjective Check List in the most recent Mental Measurements Yearbook.

This method can also be used to evaluate the course or program more directly (rather than the effects) by asking the students or participants which adjectives (or short phrases) they would choose to describe the activity:

*For me, this program has been (check as many as apply):

_____ informati[illegible]	_____ frustrating
_____ boring	_____ enjoyable
_____ motivating	_____ tedious
_____ confusing	_____ easy
_____ encouraging	_____ unorganized
_____ competitive	_____ one of the best
_____ useful	_____ a waste of time
_____ too theoretical	_____ exciting
_____ engaging	_____ too demanding
_____ tiring	_____ cooperative
_____ confidence-building	_____ challenging

B. Rank Ordering: In this method the respondent is given a small number of items (perhaps three to six), asked to compare them to one another, then decide which items are more preferable and which are less preferable. The respondent usually indicates the order of preference by writing the appropriate numbers on a questionnaire:

> *Listed below are several groups of people. Please decide which groups you would prefer to live next door to and spend a lot of time with. Then write a "1" beside the group you would most prefer, a "2" beside the one you would prefer somewhat less, and so forth, ending by writing a "5" beside the group you would least prefer:

_____ White Americans	_____ Nigerians
_____ Puerto Ricans	_____ Arabs
_____ Black Americans	_____ Japanese
_____ Chinese	_____ Latinos
_____ Chicanos	_____ Italians

By asking such questions before a course or program (perhaps during the needs assessment), teachers and trainers can gain useful information about those who would learn from them. Such questions can also become the basis for an evaluation when they are asked again at the close of the course or program and perhaps several months later. If attitudes toward some of the groups have changed during the course or program, this change may be reflected in the average ranking for each of the groups after the learning sequence (especially if the alternative groups included in the question are carefully chosen by the evaluator).

The rank ordering method can also be used quite productively to compare and evaluate different components of the course or training program. The subjects covered can be isolated and compared, or the teaching methods can be isolated and compared, or both can be combined:

*How helpful (or interesting, or valuable, or useful, or new, etc.) have the following parts of the course been for you? (Please write a "1" beside the part which was most helpful, a "2" beside the part which was less helpful, and a "3" beside the part which was least helpful):

_____ lecture on communication
_____ lecture on perception
_____ lecture on cultural adjustment

*How helpful have the following parts of the course been for you? (Please write . . .etc.):

_____ discussion of cultural conflict
_____ film on cultural conflict
_____ role playing of cultural conflict

*How helpful have the following parts of the course been for you? (Please write. . .etc.):

_____ lecture on communication
_____ film on cultural conflict
_____ simulation of cultural adjustment

The results of such questions can become valuable guidelines as the next course or training program is being designed. When the teacher, for example, then compares these results with the results from the same questionnaires administered during subsequent courses, still more valuable feedback can be obtained on the subjects and teaching methods and on the teacher himself (on the relative effectiveness of a lecture given repeatedly, for example). Furthermore, suggestive information can be obtained on the particular students in each class; (students from certain cultures may respond very positively to a given

teaching method in one class, for example, whereas students from another culture in the next class may respond to the same method very negatively).

C. Scales: A variety of scales are used in evaluation, especially where the evaluator wishes to measure perceptions and attitudes. Two basic kinds of scales will be outlined here. The first is based on lines, usually horizontal lines with a sequence of positions located along the lines:

*Please indicate how strongly you agree (SA) or disagree (SD) with each of the following statements by putting a check at the appropriate place on each line:

	SD | D | A | SA
The reasons for a culturally different person's behavior are easy to understand	
I can learn a lot from foreigners	
I really enjoy meeting Black Americans	
People in other countries would be better off if they adopted more American ways	

Although questions dealing with different subjects (self-esteem, ethnocentrism, curiosity, etc.) are usually mixed on the questionnaire, they are grouped according to subject during the design (so as to ensure adequate coverage for each subject) and during the tabulation (in order that the results can be broken down into each subject and compared).

The desired response is decided upon by the evaluator (or a small group of "objective" informed individuals). Asking the same questions (or closely comparable questions -- as determined by similar results from repeated tests with different groups) before and after the course or program will probably reflect some of the shifts in perceptions and attitudes which occurred during the activity. For purposes of tabulation, numbers can be assigned to each of the response categories (SD = -2, D = -1, A = 1, SA = 2, assuming "strongly disagree" is the least desirable response. If it is the most desirable response, then the numbers assigned would be reversed (SD = 2, D = 1, A = -1, SA = -2). An average of -1.5 then, at the beginning of the class for a given student (or for the whole class) would be a low rating. An average of 1.8 at the close of the class would suggest considerable improvement.

A second kind of scale is based on numbers. The series of numbers can include polar opposite positions (with a central position), or it can progress from a minimum to a maximum position. An example of the first, a series providing opposite alternatives, is the following:

*Please indicate how strongly you agree or disagree with the following statement by circling the appropriate number:

This training program has been valuable for me:
strongly disagree - 3 2 1 0 1 2 3 - strongly agree

In tabulating this kind of question, the numbers to the left of zero can be entered on the tabulation sheets as negative numbers. In presenting and interpreting the results, both the range of responses (from -2 to +3, for example) and the average (2.5, for example) are useful.

As mentioned above, a series of numbers can also progress from a minimum to a maximum position. Examples of a series providing progressive alternatives are the following:

*To what extent has this course been valuable for you? (Please circle the appropriate number):

not at all - 0 1 2 3 4 5 - extremely

*To what extent has this course increased your understanding of different (perhaps specified) cultures?

not at all - 0 1 2 3 4 5 - a great deal

*As a result of this course, how much increase has there been in your motivation to get to know individuals from different (perhaps specified) cultures?

none - 0 1 2 3 4 5 - a great deal

*How valuable have the following parts of this course been for you? (Please circle the appropriate number for each part):

lecture on communication	not at all - 0 1 2 3 4 5	- extremely
lecture on perception	0 1 2 3 4 5	
self assessment of skills	0 1 2 3 4 5	
inventory of prejudices	0 1 2 3 4 5	
exercise on intercultural learning	0 1 2 3 4 5	
status game	0 1 2 3 4 5	
assumptions & values checklist	0 1 2 3 4 5	
exercise on listening	0 1 2 3 4 5	
BAFA BAFA	0 1 2 3 4 5	
case study on conflict	0 1 2 3 4 5	
film on perceptions	0 1 2 3 4 5	
role playing	0 1 2 3 4 5	

Once the tabulation has been completed, the value of each part can be compared with the value of the other parts. Furthermore, by calculating the average of all of the parts, the overall value of the course or program can be estimated.

Somewhat different scales based on numbers can be designed and used when it is judged to be appropriate. They can serve both learning and evaluation purposes when the scale is built into an exercise used in the course or program. By having students participate in such an exercise early in a course, and then again (in at least the relevant parts of the exercise) at the close of the course, the teacher can trace shifts in their perceptions, for example, during the course.

A third kind of scale combines both numbers and a line on which the respondent locates his or her position. One advantage of this kind of scale is that it represents the range of alternatives both graphically and numerically, thereby making the alternatives clearer to the respondents (especially if some respondents think more easily in terms of graphics and others in terms of numbers, as well they might if they come from different cultures). The Center for Research and Education in Denver, Colorado, for example, uses the following format for some of its questions during the evaluation of its predeparture programs:

*Clarity of the Training Objectives

(circle one number)

5 The training objectives were clearly communicated and I understood them well.

4

3 I have some idea of what the training objectives were, but they were not communicated very clearly.

2

1 The training objectives were not clearly communicated and I did not understand them.

D. Written Training Exercises: As pointed out above, when the teacher or trainer uses exercises which involve the recording on paper of perceptions, attitudes, etc., these can easily become the basis for an evaluation. If the student or participant knows about the evaluation, this may, of course, alter somewhat his responses (though not necessarily reduce his learning). It is often best to make provisions whereby the results of the exercise, when given to the teacher or trainer afterwards for evaluation, are anonymous.

One example of the kind of exercise which would be suitable for the purposes of an evaluation is the "Inventory on my Prejudices" (one of many exercises included in Pusch, Multicultural Education: A Cross-Cultural Training Approach, Intercultural Network, Inc., 1980). In the exercise the student answers a series of questions regarding the frequency with which he feels or displays prejudice. By putting the responses of all the students together, the teacher can prepare an overall summary of their prejudices (as they have -- anonymously -- reported them). By doing the exercise both at the beginning of the course and then again at the end, the teacher can obtain an estimate of the changes in their prejudices during the course.

A third example is the "Profile of Attitudes and Feelings" (also in Pusch, Multicultural Education). In this exercise the student places tabs, each with a different word on it, at any point he wishes along a continuum from positive to negative, the particular place representing his feelings regarding the word on the tab. By putting the responses of all the students together, the teacher can draw up a rough profile of his students' feelings regarding, say, "foreigners," "Blacks," "my neighbors," "Chinese," etc. If this exercise is done twice, once at the beginning of the course and again at the close, an estimate can be made, for example, on how much more positively the students may now feel toward "Blacks." By combining all of the results from the second time, and comparing these with the first, the teacher can get an idea as to the amount and kind of overall shift during the course. In order to do this, the teacher would probably find it easiest to assign numbers to different points along the continuum before making a tabulation and then enter on the tabulation sheet the number for the location at which each word was placed. These would be handled, in other words, in a way similar to the scales above using lines along which the students indicate their positions. Specific ways to illustrate the results will be suggested later in the section on the appropriate formats for presenting different kinds of findings.

Using particular training exercises for learning purposes, and then using the exercise recording sheets for evaluation purposes, is one of the best ways in which learning and evaluation can be integrated.

E. Critical Incidents: The teacher or trainer can also integrate learning and evaluation by using appropriate critical incidents. These can be selected from incidents which have already been written, or which the teacher constructs, or which the students construct with the teacher's guidance. Once selected, the incidents are presented to each student individually, the purpose usually being to evaluate some of the student's intercultural skills and abilities. The primary focus is usually upon the student's ability to uncover and analyze the cultural components of a specific situation in which misunderstanding or conflict has occurred:

*(A concise description of an appropriate critical incident)

There are four major cultural factors in the incident described above. What are they?

1. _____
2. _____
3. _____
4. _____

Which one of these factors is the primary cause for the misunderstanding in the incident?

#_____.

Please explain specifically how this factor contributed to the misunderstanding:

_____
_____
_____

What <u>specifically</u> should be done to clarify this misunderstanding?

_____
_____

Given their cultural differences, what is the most significant difficulty likely to be in the future between the people involved in the incident?

_____
_____

Having decided beforehand what the best answers are to these questions, the teacher can then evaluate the adequacy of the students' answers, thereby obtaining an estimate of the student's ability to recognize, analyze, resolve and anticipate. If the same exercise is done at the close of the course with comparable critical incidents, an estimate of the increase in the student's abilities can be obtained.

Once the students or participants have each written their answers, the learning purposes of the exercise can more fully be realized if they discuss their answers with one another. Having thought through the incident individually, and having actually written down their analysis, they will probably have a much more productive discussion.

<u>F. Case Studies, Films, Photographs</u>: The purpose of case studies, films and photographs is much the same as that of critical incidents: to present to the student or participant a carefully structured, provocative, intercultural situation with a specific setting, identifiable individuals and groups, and pertinent details, to which the student can react, and on which he can sharpen his skills. The particular skills with which we are usually most concerned are his skills in isolating the cultural forces present in the situation, analyzing their effects, deciding what to do, and anticipating subsequent difficulties and possibilities.

The questions to the student after having read the case study or seen the film can be similar to those for the critical incidents. The requirement for specific responses is equally important. The teacher's decisions as to the best responses must, of course, be equally accurate, for these become the criteria in terms of which the students' responses are evaluated. The teacher should also be open to the possibility, of course, that the student may come up with some insights which the teacher had not.

As is true with critical incidents when they are used for evaluation, the student or participant usually studies the case or reflects upon the film or photographs privately, then prepares a written reaction. Having done this, discussion of the case may be very informative. After the discussion, the case may be read or the film viewed again thereby stimulating motiviation and reinforcing learning.

If the case study or film is used more strictly for purposes of evaluation, the film, say, would be shown early in the course or program and the student would answer the integrated series of questions which guide his analysis of it. There would be no discussion. At the close of the course, the film would be shown again and the questions would be answered again (discussion afterward might then be very beneficial). Seeing the film a second time at the close of the course can be rewarding for students (and, of course, teachers) because they realize how much more can be seen in the film having taken the course. New awareness and skills will be reflected in responses to the questions at the close of the course. The increase in awareness and skills can be demonstrated to the student (and perhaps to others) by contrasting these responses (privately or anonymously) to the ones written down early in the course.

A thought on who is being evaluated: we have been assuming that evaluation would focus on the individual student or participant. Given their highly individualistic orientation, this is usually the way most Americans do their evaluating. There is an alternative, however, to which some careful attention should be given. A group or a team can be evaluated. A small group, for example, would study a case, discuss it, arrive at some consensus on the answers to the questions, then record their consensus on the questionnaire. The teacher or trainer then evaluates the group's answers. This approach can strongly influence the interaction within the class or program and the cohesiveness of the group, of course, and it can increase the opportunities for learning, especially if two or more cultures are represented in the group (if, in other words, members of the group engage in intercultural interaction as they are cooperatively analyzing an intercultural interaction). Relations within the group can be further developed if, after providing the more formal questions on the questionnaire, the teacher or trainer asks the members of the group more personal questions and then discusses their responses with them. One question might be, "How would you feel if you were Lu Chen Chang in the situation we have been discussing?"

If the class or program group is divided into several sub-groups (as a large group would have to be), if each sub-group is evaluated on its combined awareness and skills, and if the evaluations are then reported to the whole group, this will, of course, further influence relations within the group. As each sub-group compares its "grade" to that of the other sub-groups, some intergroup rivalry may develop. If the teacher's or trainer's attitudes and methods are appropriate, however, intergroup competition can be more constructive than the interpersonal competition stimulated by individual evaluation. Intergroup competition can be especially constructive, and is sometimes preferable, where there are a number of cultures represented in the group. In a multicultural class or program, if the individuals are evaluated separately, the individual's isolation and inclination to return to his own culture group may be intensified.

If discussion is to take place within small groups in a multicultural class or program, and if an evaluation is to be done of these groups, each group should itself be multicultural. If each group is composed of only one culture (Chicanos in one, for example, and Anglos in another), intercultural interaction is reduced and stereotypes may be reinforced.

G. Sentence Completion: The first part of each sentence (like the method above but in a much abbreviated form) sets a certain situation or identifies particular individuals (often including the student or participant himself). The respondent then finishes the sentence in whatever way he wishes. This method can be used to estimate the present level of, and subsequent learning in, all areas: knowledge, perceptions, attitudes, skills and patterns of behavior:

*When a traditional Japanses woman attends her first American cocktail party, she is likely to feel

_____

*What Native Americans probably want most in life is

_____

When I am with a group of Mexicans I feel

_____

*When I find I am in strong disagreement with an Arab, I usually

_____

*While I am listening to a male Anglo friend who is telling me about a personal problem he has, I can show him I am actively listening to him by

_____

When the evaluator has decided specifically what he is looking for and has constructed the first part of each sentence, the specific criteria for completing each sentence can be established (against which he will compare and assess the responses of the students or participants). In addition to determining the adequacy of each response (and perhaps calculating the overall adequacy of all the responses), the pattern of the responses can be determined; this can be done by putting each response into one of several categories and then examining their

frequency. Some of the questions that may then be asked are: What kinds of responses do Blacks usually give? What kinds do Puerto Ricans usually give? What do their different responses imply for the relations between them? For my teaching of them? My evaluating of them?

If the sentences are completed by the students early in a course, and if the same (or comparable) sentences are completed at the close of the course, the teacher can demonstrate changes in the patterns of the responses and can estimate the increase in the adequacy of the responses (i.e., the increase in the students' knowledge, etc.).

H. Questions: This is the method with which people are most familiar. If the method is to be effective and the results accurate, each question must be very carefully constructed. Each must have a clear and specific purpose. The words must be carefully chosen and the order of the words must be deliberately arranged; such care is especially important in multicultural groups. The order of the questions themselves must also be thoughtfully arranged; each must build upon the ones preceding it, providing logical sequence and development throughout. Students and participants learn not only from the questions they think about but also from the order in which they think about them. Integrated questions can help integrate knowledge, integrate experience, and integrate knowledge with experience.

There are two basic kinds of questions: structured and open-ended (or free response). With regard to structured questions, the multiple choice format is the standard one; it is sometimes the best and (for the respondent, at least) often the simplest. It is normally used to test specific knowledge but it can be used more broadly. The examples below cover the areas of knowledge, attitudes, skills and patterns of behavior:

*When a White American first meets a Black American, the primary non-verbal difficulty they are likely to experience is (please circle the letter beside the correct answer):

a. personal space
b. kind of physical contact
c. eye contact.
d. none of the above would cause any difficulty

*How willing would you be to spend time assisting a new student who has just arrived from a foreign country?

_____ very willing
_____ willing
_____ not very willing
_____ not at all willing

*If you have recently moved to an Asian country and are having serious difficulty getting along with an Asian in your organization, what should you do?

a. Be polite to him and avoid him whenever possible.
b. Confront him directly and discuss the difficulties with him.
c. Express your frustrations to him, but do it in a light, humorous manner.
d. Select someone to be your interpreter and discuss the situation with him.

(These three kinds of questions, especially the second, can become the basis for more revealing evaluation by asking them before the course or program (perhaps during the needs analysis), then again at the close, and then contrasting the results.)

*During the last two months, how often have you initiated contact with a person from a different (perhaps specified) culture?

_____ as often as possible
_____ frequently
_____ sometimes
_____ seldom
_____ never

This same kind of question can then be asked again two months after the course or program; the results may suggest that some of the students or participants are initiating contact more often than before their learning sequence. It is usually better to get the same information through someone's direct observation or recording of the contacts but this, of course, is seldom possible. The learner's own reports must therefore be used and an assumption of consistency in his reporting must be made.

The same kind of information can be sought in another way during the follow-up evaluation; one example:

*We are interested in knowing whether or not this training program has encouraged you to initiate contacts with people from different cultures during the two months since the program ended. During these last two months, how many people (not from your culture) have you started a conversation with, people whom you probably would not have approached had you not participated in this program?

_____ none
_____ 1 or 2
_____ 3 or 4
_____ 5 or 6
_____ more than 6

With regard now to the second basic kind of question, the open-ended question calling for a free or essay response, this can also be used to gather information in all five areas (knowledge, perceptions, attitudes, skills and patterns of behavior). This is the kind of question used above in the examples for Critical Incidents (although structured questions can also be used, of course, with Critical Inci-

dents, as well as with case studies, films and photographs). One of the essential student or participant skills which open-ended questions can help to measure, which structured questions and other methods usually cannot, is skill in synthesizing, criticizing and applying what has been learned in the course or program.

One variable which the evaluator should keep in mind as the open-ended questions are being constructed is the extent to which each question guides and limits the student's or participant's response. More specific questions, and usually longer questions, restrict more the range of possible responses. Whether this is desirable or not depends largely upon the backgrounds and expectations of the students or participants, the purpose of the course or program and the purpose of the particular question.

Some examples of open-ended questions are the following:

*What are your two major objectives in taking this course?

1. _____
2. _____

(By asking a question regarding their objectives during the needs asssessment or early in the course, the teacher can assist the students in clarifying their objectives and assist himself in designing the course. Their objectives, once established, can then become a basis for monitoring the relevance of the course as it proceeds and for evaluating the course as it closes. The questions at the close of the course would ask the students to what extent their objectives had been met and which parts of the course contributed the most toward meeting their objectives.)

At the close of the course or training program the following kinds of questions might also be asked:

*In what specific ways has this course itself been a product of North American Mainstream culture, reflecting basic values and procedures of that culture?

_____
_____
_____

*In what specific ways has your own culture influenced your participation in this course?

_____
_____
_____

*Please list and describe briefly the seven phases of the intercultural learning process:

1. _____
2. _____
3. _____
4. _____
5. _____
6. _____
7. _____

*What, specifically, is effective feedback?

_____
_____
_____

*Please describe briefly how you gave (or could have given) verbal feedback during the course to a particular person from another culture:

_____
_____

*Describe briefly how you gave (or could have given) non-verbal feedback during the course to a particular person from another culture:

_____
_____

*"Sterotypes are necessary and very constructive in intercultural relations." Discuss this statement:

_____
_____
_____
_____
_____
_____

*Assume you are going to move to Peking next week. You will live alone and will be there two years. What specifically will you do during the first two months to orient yourself and facilitate your adjustment to life in Peking?

_____
_____
_____
_____
_____
_____

*In what specific ways is your own approach to cultural adjustment consistent with the basic values and patterns of behavior in the culture from which you come?

_____
_____
_____
_____

*What are the two most significant things which you have learned from this course?

1. _____
2. _____

In deciding upon the appropriate order for the questions, it is often best to put questions asking for an overall evaluation of the course, like the last one above, at the end (after the student or participant has thought through and evaluated the particular parts of the course or program in the preceding questions). If scales and structured questions are used, the student's thoughts and feelings about these will probably influence his responses to any open-ended questions which follow. This influence is sometimes desirable; the structured questions can prepare the student to respond to the unstructured questions. If the teacher does not want to take advantage of this influence, the open-ended questions can be put first. As mentioned above, students' responses can be even less channeled by making the first open-ended questions brief and rather general. As the teacher moves away from structured, specific questions, of course there is a greater reliance upon more "subjective" means for rating the responses, less upon "objectve" and quantifiable means.

I. Journals: Diaries and journals are the least structured method of evaluation used in intercultural courses and training programs. The instructions to the student, for example, are usually:

> *During this course, write down in a special notebook after each class your thoughts and feelings about the class, about intercultural relations, and about your participation in both. Be sure to include the most important things you are learning from the class and the things you have done (or plan to do) differently because of what you are learning.

The Journal can be read by the student only, or by the student and the teacher; the student should be told which. If the teacher is to read it, the student should also be told the teacher's purpose in reading it (interest, evaluation, discussion with the student, etc.).

If only the students are to read their Journals, frequent review of what has been written will be helpful. The reviews will provide opportunities to reinforce and consolidate their learning. Furthermore, each time they look back and see what they have learned, they will probably be more motivated when they enter class next time.

If the teacher reads the Journals and uses them simply as a basis for discussions with the students, the students' expectations will influence what is put into them. If the teacher evaluates the students' learning on the basis of the Journals, this will probably influence even more what the student puts into them. The teacher can determine to some extent the kind of influence by specifying the particular criteria to be used in doing the evaluation. Exactly what the student writes down, of course, will also be partly determined by the amount of rapport between him and the teacher. From the Journal, especially if the rapport and trust are high, the teacher will be able to draw information on changes in all five areas: the student's knowledge, perceptions, attitudes, skills and patterns of behavior.

DISCUSSION (Self-Report)

The method which combines paper and pencil procedures with discussion is interviewing. An interview schedule (containing instructions to the interviewer and the questions to be asked during the interview) must be constructed as carefully as a questionnaire. Decisions regarding the subjects to be investigated, the word order in the questions (either structured or open-ended), and the sequence of the questions are as important on schedules as they are on questionnaires. Also important is the decision regarding who is to do the interviewing; the age, sex, nationality, status and appearance of the interviewer, as well as interviewing skill, will influence considerably the results of the interview -- especially in a multicultural situation. During a course, if the students interview one another, they will probably learn a great deal (especially in a multicultural class) but the teacher may learn less. Teachers who do the interviewing have more control and more opportunity to clarify responses; also they get to know the students better. On the other hand, interviewing takes time and the teacher's presence will probably influence, and perhaps inhibit, the student's response (the presence of a peer interviewer will also influence the response, of course, but the kind of influence will usually be different from that of the teacher). Regardless of who does the interviewing, the schedule can be constructed so as to enable the teacher to evaluate knowledge, perceptions and attitudes, and obtain the student's descriptions at least of skills and patterns of behavior.

Less structured discussion can also be used in evaluation. During a training program, for example, the participants can discuss with one another what they are learning (at these points, as demonstrated in the diagram, the evaluation would monitor the participants' progress and the program's appropriateness); at the end of the program the participants can discuss with one another what they have learned (at this point the evaluation would measure the effects of the program). One advantage of this kind of discussion is that each participant learns from what the other participants have learned. Other kinds of discussion which can also be analyzed would be discussions of case studies, films and so forth.

In order to base a course or program evaluation on rather unstructured discussion, the discussion should be recorded mechanically (on audio or video tape or by a court recorder) so that the contents can be analyzed. If mechanical recording is not feasible

then an observer (or the teacher or trainer) should be appointed to take systematic notes (methods of observation will be described below). Like open-ended questions, discussion allows for some evaluation of the student's and participant's ability to synthesize, criticize and apply what has been learned.

ENACTMENT (Teacher-Report, Perhaps Peer-Report)

When using the paper and pencil and discussion methods described above, the person who reports on attitudes, what has been learned, etc., is usually the student or participant himself. Beginning with this section on enactment and in the following sections, the person who reports is often the teacher and trainer or the student's and participant's peers.

An enactment situation in the classroom, if it is carefully prepared by the teacher (or by the students themselves with the guidance of the teacher), will provide for the students an opportunity to demonstrate what they have learned, thereby reinforcing their learning and presenting to an observer (the teacher, colleague or other students) behavior which can become the basis for an evaluation. If an attempt to enact a situation reveals that students have not learned what they were expected to, not only should the student's motivation and ability be reconsidered but the design and methods of the course should also be re-evaluated; (the evaluation serves here a feedback function, suggesting points at which changes in the course should be made). If, on the other hand, the enactment reveals that the students have learned what they were supposed to, the teacher can be reasonably confident that the course is on target.

The variety of situations which can be enacted in courses and programs is wide: role plays (and reverse role plays), simulations, skits, plays, critical incidents (where, before discussing the incident, it is acted out), debates (where, for example, the students must argue a position contrary to their own and do so on the basis of values very different from their own), or common tasks for the class or small groups within the class (which require that the students plan, decide and cooperate with others from different cultures).

If any of these situations (a particular role play, for example) is enacted early in the course and then again at the close, the students can more clearly recognize and appreciate how much they have learned, and the teacher, by contrasting the students' performance during the first with their performances during the second, will be able to estimate the degree of improvement.

The method of enactment makes possible, more than the other methods above, the direct and precise observation and analysis of the student's interactive skills and patterns of behavior. Data on these are usually gathered through the teacher's and sometimes the students' observations (methods of observation are described below). In deciding what to enact, and in deciding exactly what to observe, the teacher or trainer must, of course, have established specific criteria for effective performance in each situation. These criteria must also be consistent with (in fact are best derived from) the goals and objectives of the course or training program.

PRODUCTION (Teacher-Report, Perhaps Peer-Report)

This method involves the construction of a tangible product. In intercultural courses, for example, the construction can be undertaken by the individual student or by a small group. Possible kinds of products are limited only by the goals of the course and the imagination of the teacher and students. The alternatives include photographs, poems, drawings, sculptures, short stories, essays, conceptual models, case studies, critical incidents, scripts, video tapes, films and so forth.

What exactly is to be evaluated? In the previous section where we described situations which can be enacted, the primary focus was upon the process, usually the process of interaction among the students. Here, however, because there is a specific product, either the product or the process can be evaluated -- or both. (In deciding between them, it is well to bear in mind that, contrary to some American assumptions, it is possible to have a splendid process and a worthless product; furthermore, an unacceptable process can result in a superior product.) In multicultural education priority is often given to the process, but the teacher and the students often learn much if some attention is given to both.

If it is the product which is to be evaluated, anticipation of the evaluation can increase the motivation of an individual or the consolidation of a group (which, of course, will influence the process). From an analysis of the products, depending upon what kinds of products they are, some inferences can usually be drawn regarding all areas: knowledge, perceptions, attitudes, skills and patterns of behavior. Some estimates of the last two, for example (the student's intercultural skills and patterns of behavior), can be made through an analysis of the student's portrayal of these in the characters he creates in his short story or script.

From an analysis of the process, sometimes regardless of the product, some inferences can be drawn, again, regarding all five areas. As is true during the enactment of a situation, the construction of a product by a multicultural group makes more possible than do most other methods the direct and precise observaton of the student's or the participant's interactive skills and patterns of behavior.

As was also true during the enactment of a situation (as well as during an unstructured discussion), if it is the process which is to be evaluated, some means of recording must be provided. Because much that happens during the construction of a product is not verbal, video tape is usually preferable to other mechanical means. The alternative is systematic observation and notation by the teacher, selected colleagues or other students; some methods of observation are described in the next section.

If the individual student or small group constructs a product (a conceptual model or a critical incident with an analysis, for example) early in the course, and then constructs the same kind of product toward the end of the course, the teacher and the students can compare the content of the products and the quality of the interaction

at the two points. They may also want to compare the relative importance of the process as opposed to the product (the process may have become more important to the students during the course).

OBSERVATION (Teacher-Report, Perhaps Peer-Report)

Observation is important to a number of other methods described above, especially unstructured discussion, enactment and production. As mentioned earlier, observation during intercultural courses, for example, may be undertaken by the teacher, by one or more of his colleagues or by the students (who would learn valuable observation and intercultural analysis skills through the exercise -- especially if they compare and discuss their observations afterward).

The presence and purposes of the observer (which should be explained to the students) will probably influence the process and the behavior of the people being observed. The influence of a teacher will probably be different from the influence of a student. The influence of five students will probably be greater than the influence of one. The kind of influence which is likely to be exerted, and the effect of this influence upon the results of the observation, must be considered by the teacher as the observer is selected, instructed and introduced.

Observation is usually more penetrating and more useful for evaluation if it done systematically. As with other methods, the teacher and the trainer must first decide exactly what is being sought. All five areas can be investigated. When it is knowledge, perceptions and attitudes being evaluated, the teacher must specify what forms of behavior will probably be associated with each ("If the student has X piece of knowledge, he is likely to behave in Y and Z ways"). Therefore, when certain forms of behavior are actually observed, inferences can be made back to the knowledge, perceptions or attitudes which the behavior presumably represents.

Observation is especially appropriate for the evaluation of skills and patterns of behavior. Fewer intermediate steps are necessary than for the other areas. The best basis for an evaluation of these two areas are the behavioral objectives of a course or program; ("By the close of this program, the participant will be taking such and so actions in this specified situation"). It is exactly these actions (and consequent interaction) that the observer will look for.

In order to ensure that the observing itself is done systematically, and that the target patterns of behavior are actually recognized and recorded, it is helpful for the observer to have in front of him a list of the specific patterns (an observation schedule). The schedule should be set up in such a way that a given action or interaction can be quickly recorded by simply making a mark on the sheet. The categories which can be marked for a given individual action can include the frequency and appropriateness of the action. Categories for a given interaction can include the frequency, quality (perhaps constructiveness) and intensity of the interaction. Room should also be provided for other kinds of notation which seem pertinent to the observer and which may be important in interpreting the results. (See the next section on tabulation for a sample of an observation schedule.)

An initial alternative, of course, is to use a mechanical means of recording. Notations are then made on the schedule while listening to, looking at or reading a transcript of the interaction. The advantage here, of course, is that the teacher or trainer can back up and listen again to sections or go through the whole recording again if he wishes. In deciding whether to use an observer or a mechanical recording device, one of the considerations is that the device, too, will probably influence the process being recorded, and it will probably do so in ways different from (though sometimes preferable to) an observer.

The performance of each individual, the interaction between two individuals, and the overall process of the group -- all three can be recorded in these ways and then evaluated. If all three are to be evaluated, it is usually best to have at least one observer for each (with a separate and appropriate schedule for each) in order to allow adequate concentration.

Once the actions, interactions or group processes have been recorded, inferences can be drawn regarding the students' or participants' knowledge, perceptions and attitudes, and estimates can be made regarding the extent to which desired skills and patterns of behavior have been developed.

UNOBTRUSIVE MEASURES (Teacher-Report)

All of the methods which we have considered above require the student and participant to respond to what the teacher or trainer plans and prepares: writing, choosing, discussing, enacting, constructing, observing and recording. These methods enable the teacher and trainer to maintain considerable control and to conduct systematic, thorough evaluations.

There are three basic problems, however, with these methods. The first has to do with feasibility. In some situations it is simply not possible for the evaluator to have this much control. For the teacher or trainer, however, it usually is possible; they do have some jurisdiction over what the people involved do, how they do it and when they do it.

The second problem, a more serious one for the teacher or trainer, has to do with artificiality. Reading a critical incident is not the same as causing it. Role playing or simulating culture conflict for a few hours probably evokes some different responses than living in the midst of the conflict.

The third problem, which may be still more serious, is the possibility of "contamination." In doing an evaluation, it is difficult to determine exactly what the student or participant is responding to. The student may be responding to the teacher, the classroom setting, the methods of measurement, or a combination of these -- as well as to the subjects the teacher wants to investigate. The results may reveal to the teacher more about the effects of his methods of measurement than the effects of his course.

The problems of artificiality and contamination can he dealt with in two ways. First, the teacher and the trainer can anticipate and minimize in a number of ways the effects of extraneous factors. Some of these ways have been suggested in the descriptions of the methods above (deciding carefully who is to record observations and how they are to do it, for example, and minimizing the discrepancy between the methods used and those to which the student or participant is accustomed). What extraneous effects remain can be somewhat clarified and compensated for in the interpretation of the results.

The second way to deal with these problems is to use methods which avoid them, methods which in no way interfere with what the learner would normally be doing, methods of which the learner is completey unaware. Much attention has been given such "unobtrusive" methods during recent years. They usually focus upon the learner's behavior and patterns of behavior; they are therefore most appropriate to a course or program with behavioral objectives and should be consistent with these objectives.

If unobtrusive measures are to be used, the teacher might look at such things as library records (which students check out which titles for how long), attendance records for the class and how these change during the course, participation in voluntary activities (multicultural events, clubs, orientations for foreign students, etc.) and how this changes before and after the course, the number of times disciplinary action is required and the kind of action required, the composition of cluster groups during free time, enrollment in subsequent courses also related to multicultural education, the degree of cultural awareness represented in a small school newspaper, the kinds of summer jobs taken and careers chosen, and so forth.

Coming up with all the alternative indicators of learning and indicators of other effects of a course or program which can be unobtrusively measured requires much thought and imagination. Actually studying the indicators requires a lot of planning. Because it often means going through other people's records and files, it requires time and patience; if the records have been kept at all they are usually designed for other purposes, they are often difficult to find and some of the critical ones have inevitably been lost. Patience is also called for on the part of the other person whose files are being searched (who usually has no idea what the evaluator is after). It is therefore especially important here, as with every method, that the evaluator know beforehand exactly what he or she is looking for.

When using unobtrusive measures, therefore, suitable data are often difficult to come by. Also, once they have been collected, inferences regarding course or program effects and learner characteristics are difficult to make. When a teacher finds, for example, that 30% of his students go on to careers in multicultural education, he may conclude that one reason they do is his excellent course. In fact, however, the reason most of them took his course may have been their intention, previously formed, to pursue such a career. They may have continued on into this career <u>in spite of</u> the course. Perhaps without the course another 10% of the students would have pursued such a career.

Unobtrusive measures, therefore, although they can have some advantages, can also have some serious disadvantages. The other methods described above, the more obtrusive ones, although they do present several problems, have some definite advantages. One major advantage possible in an integrated course or program, is that their effects upon the students can be very constructive and therefore very desirable; these methods, in fact, can be so selected and so designed as to have particular effects, effects which further realize the objectives of the course.

The conclusion is clear: because of their different advantages and disadvantages, the teacher and trainer should use, whenever possible, both kinds of methods, then see to what extent (and at what points) they corroborate each other.

6

HOW THE TABULATION OF THE RESULTS CAN BE ENJOYABLE

Many teachers and trainers who could do evaluation, and learn a lot from it, don't. The reasons are often that they do not know how to do it and they are fearful of the time it may take. The preceding sections explain how an evaluation can be done. If the teacher and trainer follows the procedures outlined, the amount of time required can be minimized -- including the time required for tabulation. When the evaluator starts off knowing exactly what is wanted, then selects the methods accordingly and constructs the questionnaires and schedules accordingly, the tabulation follows easily. In other words, if the data collection is focused and systematic, the tabulation can be done very efficiently. And enjoyably.

Seeing the results emerge can be exciting. One starts with definite questions to which answers are really wanted and then goes step by step through the process of getting them. They emerge ready for incorporation into the final format and for interpretation. For many people, doing a tabulation is more enjoyable (and sometimes more efficient) if it is done with someone else. One person can read the responses to the structured questions, for example, while the other records them on the tabulation sheets; after awhile, they can switch.

The sheets on which the tabulation is done can be designed in a variety of ways. Teachers and trainers should set up those which are quickest and simplest for them. Some possible designs are suggested below.

CHECKLISTS

ANGLO PERCEPTIONS OF CHICANOS

# of students: 32

		total responses	% of total students
sincere	/////	5	16
aggressive	//////////////////////	22	69
etc.			

If this kind of question is asked after a course as well as before the course, the number of Anglo students who perceive Chicanos as being sincere, for example, may have increased considerably.

SCALES

	0	1	2	3	4	5	average rating
Question #1			/	//	////	/////	4.1
2	/	////	////	//			1.7
3							2.9

In order to figure the average rating for each question, multiply the number of responses in each column times the rating number at the top of the column, then add these products together, then divide by the total number of responses for that question; with a little practice it goes very quickly. For Question #1:

$$(1x2) + (2x3) + (9x4) + (5x5) =$$
$$2 + 6 + 36 + 25 = 69$$

This is then divided by the total number of responses (17) to the question:

$$69 - 17 = 4.1$$

After the rating for each question has been determined, they can be compared with one another. Also, the ratings for all the questions of a particular kind (those evaluating parts of the course, for example) can be added up, then divided by the number of questions, thus providing an average rating for the whole course. The ratings for each part of the course and for the whole course become especially interesting when compared with the ratings for the next course.

MULTIPLE CHOICE QUESTIONS

In multiple choice questions the alternatives from which the respondents choose are not numbers, therefore an average rating is not possible; instead, the percentage of respondents who chose each alternative is determined:

QUESTION # 4

		total responses	% of total students
as often as possible	/	1	1
frequently	//	2	12
sometimes	///	3	18
seldom	///	3	18
never	////////	8	47

This question for an intercultural course might ask, "How often would you be willing to spend time assisting a new student who has just arrived from a foreign country?" If the question were asked before the course and the above responses were given, it could be asked again at the close of the course, by which time the students might be willing to spend much more time with new foreign students.

ESSAY RESPONSES AND SHORT ANSWERS

The evaluation and tabulation of essay or free responses, given to incomplete sentences and open-ended questions, is sometimes more difficult and less precise. The several ways of doing it divide into two basic approaches: those that compare the students' or participants responses to one another and those that compare their responses to a general standard or specific set of criteria. When using the first approach, the teacher (or his willing and perhaps more "objective" colleague) considers all of the responses to a given question, then decides which responses are better and which are worse than the others. This can be done by arranging them on a table or in a pile such that the best response is on the top and the poorest on the bottom. The teacher decides what he wants the highest and lowest ratings to be, then assigns them to the top and bottom responses and distributes the ratings in between to the rest of the responses. Having done this the teacher will have ranked and rated the students' responses, but not the course.

When using the second approach, one alternative the teacher or trainer has is to keep in mind a general notion as to what an excellent response would be, then rate each response according to how closely it approximates his standard of excellence. The rating suggests how much a given student has learned on that point; an average of all the students' ratings suggests how effectively the course has

covered that point. If all of the responses are poor in terms of the standard, and assuming the question was clear, perhaps the teacher should go into the point more extensively or use different methods to convey it.

Another alternative: the teacher or trainer can more easily and more precisely quantify essay responses by establishing specific criteria for each set of responses to a given question. He can say, for example, that if the student refers to such and so, which is most important, he will get the highest score, perhaps 20. If the student deals with another particular point, which is also important but less so, he will get a score of 15, and so forth.

A variation on the above: the teacher establishes that four points (or five, or two, etc.) should be covered (or four particular attitudes revealed, etc.) in an excellent essay response. If the student covers all four, he gets a 20; if he covers three, he gets a 15, and so forth. The tabulation sheet for this kind of rating procedure could be the following (the numbers entered are each student's score; the scores assume each response should contain five points, four points, two points or one point respectively, and that the maximum score for each response is 20):

QUESTIONS #7, 8, 9 and 14

	#7	#8	#9	#14	student's average
Mary	20	15	0	20	14
Juan	8	5	0	0	3
Aaron, etc.					
CLASS AVERAGE	14	10	0	10	9

The tabulation sheet therefore shows each student's average score on all of the essay responses written; it also shows the class' average score on each question and (in the box in the lower right hand corner) the class' average score on all the questions; it is these scores which take on more meaning when obtained before and at the close of the course or program, and then compared, and when compared from one year to the next.

A third alternative: the teacher or trainer can establish very specifically what he or she is looking for, as above, but then simply record whether or not it appears in the essay rather than deciding upon a score. This method is valuable when the teacher is more interested in discerning patterns in the students' responses than rating their responses. This kind of analysis is a form of observation (what is being observed here is an essay passage rather than an actual behavioral sequence). The tab sheet, therefore, is similar to an observation schedule.

In setting up this kind of tabulation sheet, the teacher or trainer decides specifically what will be looked for (a particular attitude, for instance), then lists these on the sheet. Next, all of the phrases which would indicate each particular attitude are written down. As the responses are read, marks are made on the sheet whenever the kinds of phrases and attitudes sought are encountered. The frequency of each is then calculated.

Responses pertaining to all five areas can be recorded; patterns of behavior, of course, are those reported by the student or participant. A comprehensive tab sheet of this kind might look like the one on the following page.

	KNOWLEDGE				PERCEPTIONS				ATTITUDES								SKILLS				BEHAVIOR							
					Positive			Negative	Constructive				Obstructive								Constructive				Obstructive			
	about Anglo American culture	about Black American culture	patterns of cultural adjustment	etc.	of self	of Japanese	etc.	(same)	curiosity	motivation	confidence	etc.	exploitation	suspicion	rejection	etc.	cultural analysis	interaction analysis	conflict clarification	etc.	listening	helping	learning	etc.	judging	dominating	ridiculing	etc.
Juan	✓	✓									✓		✓													✓	✓	
Mary					✓					✓									✓		✓	✓	✓					
Aaron			✓			✓			✓						✓		✓	✓	✓						✓			
etc.																												
FREQUENCY																												
% OF TOTAL STUDENTS																												
AVERAGE %																												

From the patterns which emerge during this kind of analysis, the teacher can learn much about the students and the course. If the analysis is done at the close of the course, the results may suggest which areas have been neglected (or those areas to which the students have been least receptive). If the analysis is done both before and after, the changes in the patterns (perhaps as a result of the course) can be traced. If then the same analysis is done of the students' essay responses in subsequent courses, changes in the patterns from one course to the next can be traced. The teacher may find, for example, that each year for the last three years the percentage of students who manifest constructive attitudes has increased sharply, while the percentage who manifest specific knowledge has decreased sharply.

In attempting to account for these changes, the teacher will begin an interpretation which will consider changes in the course and methods during these years, changes in the backgrounds of the students enrolling in his course, and so forth. Having done this, there will be a very sound basis for re-designing the course. By sharing his findings with his students, colleagues and administrators, the teacher can help others clarify their own understanding of, and planning for, multicultural education courses. Whether or not these people will listen and assimilate the information will depend in large part upon the format in which the findings are presented.

7

APPROPRIATE FORMAT FOR THE FINDINGS

Once the tab sheets are completed, the teacher or trainer can decide which is the best format in which to display the findings. Sometimes the tab sheet itself is best. Often, however, it contains more data than are needed. Furthermore, if the findings are to be clear and easily grasped, a format is usually needed which is more visually representative, one which presents the data in lines, bars, graphs and contrasting locations on the page. Such a simplified, easily assimilated format is especially important when the results are to be presented to someone else (the students or participants, administrators or managers, clients, etc.).

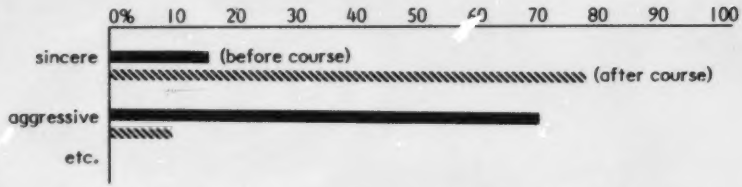
Some formats tend to be more suitable for the results obtained through certain methods, other formats for the results from other methods. Several of these formats are illustrated below. Often, however, a given set of results can be displayed in a variety of formats. The teacher or trainer can therefore select the one which is most appropriate to him, his class, his methods and his audience. Regardless of the format chosen, it is usually best to present the question or the instructions in order that the person looking at the results can know exactly to what the students or participants were responding.

CHECK LISTS

ANGLO PERCEPTIONS OF CHICANOS

number of respondents: 32

15. Generally speaking, Chicanos are (please check as many as apply):



When using this method and format, the teacher can estimate (or better, have the students judge -- after responding to the question) which adjectives (or short phrases) are positive, which negative and which neutral or mixed. In the graph, then, the positive adjectives can be displayed separately; an increase in most or all of the positive ones can be demonstrated, as perhaps can a decrease in most or all of the negative ones.

The results of check lists can also be clearly displayed by showing them in rank order with the number of respondents choosing each alternative shown as a percentage of the total number of respondents:

PARTICIPANTS' DESCRIPTION OF PROGRAM

number of respondents: 32

17. For me, this program has been (check as many as apply):

82%	exciting
80	engaging
76	motivating
68	one of the best
65	challenging
65	informative
61	useful
35	frustrating
22	tiring
14	too theoretical
9	confusing
0	tedious
0	boring
	etc.

When changes are made in the program the next time it is offered, the trainer can ask this question again and see if there are any changes in these responses; it may be found, for example, that in describing the revised program, 14% of the participants found it exciting and 82% found it too theoretical.

SCALES

Methods which provide lines on which the student or participant locates his or her position can provide very interesting findings. These can be illustrated in a variety of ways, one of which is the following (this assumes the positions have been converted into numbers during tabulation):

ETHNOCENTRISM, CURIOSITY AND MOTIVATION

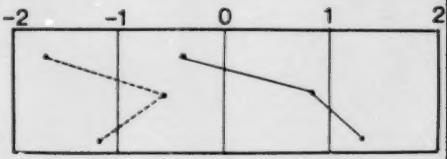
number of respondents: 32
(-2 = very unacceptable; 2 = very acceptable)

Ethnocentrism

questions 1, 4, 7 (ethnocentrism)

3, 5, 9 (curiosity)

2, 6, 8 (motivation)



--------- pre-course
_____ post-course

Scales from which the student or participant selects a number can yield equally interesting results. One format appropriate for them is the following:

EVALUATION OF COURSE GOALS

number of r spondents: 33

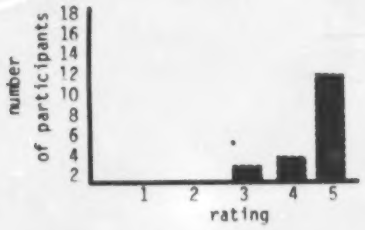
18. As a result of this course, to what extent do you now recognize more clearly the influence of your own culture upon your particular values, goals, expectations and patterns of relating to other people?
(not at all - 0 1 2 3 4 5 - much more clearly)

AVERAGE: 4.1 RANGE: 1-5

The results of an evaluation are often more quickly understood--and more easily remembered--if they are represented by both graphics and numbers. The following format is one of several used by the Center for Research and Education in Denver, Colorado to present the evaluation of its training programs to its clients:

CLARITY OF THE TRAINING OBJECTIVES

5 = clearly communicated, I understood them well
1 = not clearly communicated. I did not understand them



When several short items have been rated with a numerical scale, the results can be clearly displayed by rank ordering them. These results, of course, are extremely useful in designing the next course or training program:

EVALUATION OF COURSE COMPONENTS

4.9 exercise on intercultural learning
4.7 film on perceptions
4.7 lecture on communication
4.3 exercise on listening
3.8 case study on conflict
3.1 role playing
2.0 self assessment of skills
1.8 status game
1.0 lecture on perception

An average of all of these would give the teacher an overall evaluation of the course.

ESSAY RESPONSES AND OBSERVATIONS

With regard to essay responses (elicited through critical incidents, case studies, films, photographs, sentence completion, open-ended questions and Journals), when these have been analyzed by means of tabulation sheets, the results can be presented in graphs, tables or other formats illustrated above. These same formats can also be used for the results from methods involving observation schedules and other systematic, quantifiable means of recording (discussion, enactment, production and unobtrusive measures).

When the tabulation of essay responses is not feasible or desired, the responses can be presented in a narrative which carefully summarizes the answers to each question and points out the frequency of each kind of response. This narrative can be more clear and engaging if a few representative responses are actually quoted.

8

INTERPRETATION OF THE FINDINGS

The teacher knows better than anyone else the purpose and content of his course, the backgrounds of the students in it and the appropriate methods of evaluation. The teacher, therefore, is in the best position to explain the meaning of his findings. For similar reasons, the trainer is in the best position to explain the significance of his findings. It is the teacher's and the trainer's obligation (and opportunity) to see that the results are interpreted responsibly.

The best time to do the interpretation is immediately after the tabulation. In fact, the evaluator will probably find irresistible the making of interpretative notes in the midst of the tabulation because questions, ideas and possibilities inevitably occur to an alert tabulator.

When the results of an evaluation are going to be presented to someone other than the evaluator, it is helpful to the other person if the interpretation begins with a brief description of the purposes and methods of the evaluation, the purposes and methods of the course or program in which it was done and the backgrounds of the students or participants involved.

The next step, whether the interpretation is being done for oneself alone or for others too, is to study all of the results and select those which are most important -- important in that they are most unexpected, revealing or useful. Care must be taken in selecting these results and caution must be exercised in interpreting them. What seems to have been a change in perceptions or attitudes, for example, may not have been. In the absence of statistical tests of significance, small changes should not be the basis for conclusions.

An increase of 10% in positive perceptions after a course, for example, probably demonstrates little one way or another about the course. An increase of 30% or more would probably be more suggestive and should be taken more seriously.

Caution is also called for in interpretation because the methods of social science are not precise, nor are the numbers which result; the numbers often suggest an accuracy which they do not possess. It is therefore with substantial changes, basic contradictions and consistent patterns over time that the interpreter is concerned.

Having selected the most important results, the evaluator tries to account for them. If a change in a particular perception is indicated, for example, everything to which the student has been exposed which might have brought that change about is considered: a particular series of exercises in this (or another) course, a major unplanned event in the course, the instructor's teaching style, the students' participation in certain methods of evaluation, relevant experiences the student or his friends have had outside the school in the community and their families and so forth. The teacher then reviews all the other results and recalls his own observations of the class in order to pick up any further clues. The most plausible explanation for the change can be identified and then incorporated into the interpretation. Again, caution is appropriate: cause and effect relationships are difficult to establish in the social sciences. (The connection between cause and effect is sometimes especially tenuous when using unobtrusive measures.)

The final step in the interpretation: the teacher or the trainer answers the basic question, So what? Given these data, given these important findings, what specifically should I do? What changes should be made in the course or program? What should be kept, what modified, what eliminated? What should be created? It is here, at this point, that evaluation can begin to make a profound difference in the nature and depth of intercultural learning offered to students and program participants.

Responsible interpretation, then, calls for thinking carefully into the results and making judgments which are backed up by an explicit rationale. Repetition is also extremely important. As the evaluation of the course with different students in different years is conducted and carefully interpreted, fundamental patterns begin to emerge. The teacher's and trainer's knowledge of these patterns becomes the basis for more accurate, more confident interpretation -- and for more effective multicultural education and intercultural training.

9

CONCLUSIONS

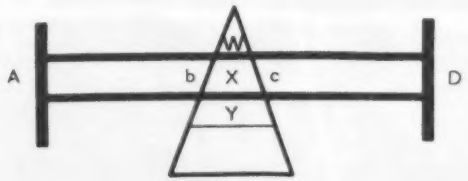
Particular perspectives on learning and methods used by teachers and trainers usually develop within, and become increasingly reinforcing to, a single culture. These perspectives and methods are therefore uniquely suited to students and participants who were themselves brought up in that culture. They are usually inappropriate, and can even be damaging, in multicultural classrooms and training rooms. Teachers in multicultural education and trainers in intercultural programs therefore have a responsibility to explore new perspectives and create alternative teaching styles and methods. In the midst of their experiments; they must measure the effects of each. The results of their measurement can then become the basis for selecting those approaches which are most effective. Having more fully and confidently implemented these approaches, further evaluation becomes the basis for improving and refining them.

As we do our designing, measuring and interpreting, it is well to recognize that the assumptions and methods of evaluation have themselves developed within particular -- for the most part Western -- cultures. Many assumptions and methods are distinctly American. This does not mean we should not do evaluation within multicultural groups. It means that we should do it as carefully as possible, being sensitive to the diverse backgrounds of our students and participants, and working toward those perspectives and methods which will respect and accommodate contrasting, even contradictory, expectations and preferences.

In addition to being limited by our culture, our methods are limited by our subject -- learning. Because of the complexity of any learning, and because it occurs within a human being (not a machine), we can measure only some of what some of our students and participants learn. If, for example, we would like to measure the amount of change

in their knowledge, perceptions, attitudes, skills or patterns of behavior, we can measure only part of that change. Looking at the diagram "Measuring Amount of Change," let's assume that their knowledge has increased from point A to point D. We then do our evaluation. The area in the triangle represents the area

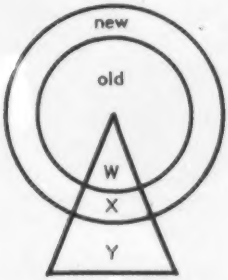
MEASURING AMOUNT OF CHANGE



of knowledge revealed by the evaluation. We can see that the evaluation will reveal only the increase from b to c. As our selection of methods becomes more informed, and as we improve upon these methods, this area can be increased. As both our methods and our interpretations become more precise, we can be increasingly confident that it is in fact X that we are describing, rather than W (old knowledge) or Y.

In evaluating the effects of courses and training programs, as explained earlier, it is often desirable to distinguish between old knowledge or previous perceptions (what the student brings to the course) and new knowledge or altered perceptions (what the student takes from the course, presumably as a result of having been in it). The means for making this evaluation were explained in the sections above. The complexity of our subject however, and the lack of precision in our instruments continue to limit what we can discover. Looking at the next diagram, "Differentiating Old from New", we can

DIFFERENTIATING OLD FROM NEW



see that our evaluation can distinguish between old and new, but it reveals only parts of each. In our methods, in the formats for displaying the results, and in our interpretations, we must be careful that we measure and demonstrate the differences between W and X. As above, we must also be careful to differentiate both of these target areas from Y.

To whom should we show our results and interpretations? The individuals or organizations for whom we may have done the evaluation, of course, will want to see them. Our colleagues will probably be very interested in seeing and discussing them. And our students and participants would probably appreciate seeing them. In fact, some evaluators believe that we are ethically responsible to our students and participants to show them the results of our -- and their -- efforts. Because they cooperated in the evaluation, and because they were themselves the ones studied, they have a right to see and discuss the results.

There are additional reasons for going over the results with our students and participants. In doing so, they learn more about the topics of the course or program, more about their teacher or trainer and, especially, more about themselves. They also learn more, sometimes a great deal more, about evaluation. This is constructive, as mentioned earlier, in that they will then be better able to continue some self-evaluation on their own after the course or training program has been completed.

Learning about the purposes and procedures of evaluation, and especially about its capacity to provide very revealing insights, is constructive for a further reason. The men and women now learning from us will soon be the ones making the decisions and designing the courses and programs which determine the quality of, and the support for, multicultural education and intercultural training. In learning more about evaluation, which is essential for strengthening quality and demonstrating consequences, our students and participants will be able to contribute substantially to our common concern.

RESOURCES

ORGANIZATIONS

Center for Research and Education, 1800 Pontiac Street, Denver, Colorado 80220

Evaluation Associates, Box 244, 740 East Main Street, Pinckney, Michigan 48169

Institute for Behavioral Sciences, 250 Ward Street, Suite 226, Honolulu, Hawaii 96814

Intercultural Network, Inc., 906 North Spring Avenue, LaGrange Park, Illinois 60525

International Resource Development, Inc., P. O. Box 721, LaGrange, Illinois 60525

Sage Publications, Inc., P.O. Box 5024, Beverly Hills, California 90210

Systran, 70 W. Hubbard Street, Chicago, Illinois 60610

BOOKS

Abt, Clark C., ed. The Evaluation of Social Programs. Beverly Hills, California: Sage Publications, 1977, 504 pages, $25.00

Cook, Thomas D. and Charles S. Reichardt, eds. Qualitative and Quantitative Methods in Evaluation Research. Beverly Hills, California: Sage Publications, 1979, 160 pages, $12.50 hardcover, $5.95 softcover

Handbook of Evaluation Research. Beverly Hills, California: Sage Publications, two volumes, 1,400 pages, $50.00

Improving Cross-Cultural Training and Measurement of Cross-Cultural Learning. Denver, Colorado; Center for Research and Education, 1973, $6.00

Improving the Evaluation of Peace Corps Training Activities. Denver, Colorado: Center for Research and Education, 1973, $3.00

Rich, Robert F., ed. Translating Evaluation into Policy. Beverly Hills, California: Sage Publications, 1979, 160 pages, $12.50 hardcover, $5.95 softcover

Schulberg, Herbert C. and Jeanette M. Jerrell, eds. The Evaluator and Management. Beverly Hills, California: Sage Publications, 1979, 159 pages, $12.50 hardcover, $5.95 softcover

JOURNALS

Evaluation Review. Beverly Hills, California: Sage Publications, six issues per year, one year $40.00 Institutional, $20.00 Individual

Evaluation Studies: Review Annual. Beverly Hills, California: Sage Publications, 1976, 1977, 1978, 1979

OTHER PUBLICATIONS OF THE INTERCULTURAL NETWORK

MANUAL FOR MULTI-CULTURAL AND ETHNIC STUDIES
by Henry Ferguson

MULTICULTURAL EDUCATION: A CROSS-CULTURAL TRAINING APPROACH
Margaret D. Pusch, Editor

THE EVALUATION HANDBOOK FOR CROSS-CULTURAL TRAINING AND MULTICULTURAL EDUCATION
By George Renwick

SURVIVAL KIT FOR OVERSEAS LIVING
by L. Robert Kohls

MANAGING CULTURAL DIFFERENCES
by Philip R. Harris and Robert T. Moran

INTERNATIONAL AND INTERCULTURAL COMMUNICATION ANNUAL, VOLUME V, 1979;
A journal of the Speech Communication Association
Nemi C. Jain, Editor

HELPING ACROSS CULTURES
Gordon L. Lippitt and David S. Hoopes, Editors

LIVING IN THE U.S.A.
By Alison Raymond Lanier

RED DUST ON THE GREEN LEAVES
by John Gay

BOOKS PUBLISHED AND DISTRIBUTED BY THE NETWORK
FOR THE SOCIETY FOR INTERCULTURAL EDUCATION,
TRAINING AND RESEARCH

AMERICAN CULTURAL PATTERNS: A CROSS-CULTURAL PERSPECTIVE
by Edward C. Stewart

TEACHING INTERCULTURAL COMMUNICATION: CONCEPTS AND COURSES
David S. Hoopes, Editor

A MANUAL OF STRUCTURED EXPERIENCES FOR CROSS-CULTURAL LEARNING
William W. Weeks, Paul B. Pederson and Richard W. Brislin, Editors

OVERVIEW OF INTERCULTURAL EDUCATION, TRAINING AND RESEARCH, Three Volumes
David S. Hoopes, Paul B. Pederson and George W. Renwick, Editors

INTERCULTURAL SOURCEBOOK: CROSS-CULTURAL TRAINING METHODOLOGIES
David S. Hoopes and Paul Ventura, Editors

INTERCULTURAL THEORY AND PRACTICE
William G. Davey, Editor

For further information write to: Intercultural Network, Inc.
906 N. Spring Avenue
La Grange Park, Illinois 60525